INSTITUT TEKNOLOGI TELKOM

ICTel
International Conference on Telecommunications

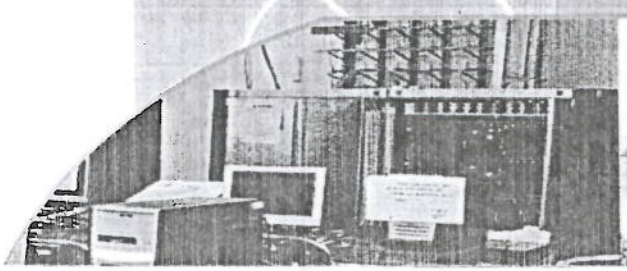
B4

Grand Preanger Hotel
Bandung, 19th-21st August 2008

19th-21st August 2008

PROCEEDINGS

International Conference on Telecommunication (ICTEL 2008)



ISSN 1858-2982

Table of Content

DATA MINING DESIGN (CASE STUDY: ANALYTICAL CRM PT. DOS NI ROHA)

Suparto Darudiato[1], Michael Maramis[2]
[1,2]Binus University
[1]supartod@binus.edu, [2]mmaramis@binus.edu

Abstract

The purpose of this research is to analyze and to design a data mining model that can be used to reveal the information of hidden customer buying behaviour in the database PT Dos Ni Roha, which involves in distribution industry. The data mining algorithm that will be used is association rules algorithm. In this research, the methodology involves gathering data, analyzing and designing methods. Gathering data method is done by field and literature study. Analyzing data consists of Porter's five forces model, issues priority matrix, SWOT, EFAS, IFAS, goal – strategy, CSF, KPI, problem, and information needs analysis. The designing method refers to the approach of Berry and Linoff. The result of this research is a data mining model that realized into a web-based application that can be used by the management to gain the information of customer buying patterns and products recommendations to help in the process of making a product penetration strategy.

Keywords: Analytical Customer Relationship Management (CRM), Data Mining, Mining Model, Association Rules

1. Introduction

In order to anticipate the changing environment, from product-focused industry to customer-focused industry, which occurs so fast, many enterprises has realized that serving existing customer properly is one of company's main sustainable revenue and growth centre. This effort gives the ability to enterprises to be standing still in a changing environment which is being tighter. A big challenge for enterprises is to comprehend customers' needs and to respond to them.

Meanwhile, data mining can detect and extract patterns which exist in enterprises databases but the management does not know, as useful information. In this term, data mining can help enterprises to extract hidden patterns of customer behaviour. These patterns will be used in the decision-making, as an action to react to the available information.

To obtain that information, enterprises need a good data mining design to extract it from abundant data. Ti\his information can assist the management to make a good-critical decision about customers.

By calculating and understanding enterprises needs, the designing-data mining is a model that can reveal information about customers buying - behaviour based on the recorded customers transactions and display it in comprehensible way. Developing other models can help enterprises to gain some knowledge, in order to plan sales strategies, such as bundling campaign, etc., to penetration efforts of principals' products.

2. Data Mining

Data mining is a process to find patterns in data. This process can be automatic or semi-automatic. Found patterns must have meanings and give some advantages, sort of economic advantages [8, 5]. Meanwhile according to [5, 7], data mining is an exploration and analysis of abundant data to find meaningful rules and patterns. Data mining can be classified into two types, directed and undirected. Directed data mining can reveal patterns of fore-chosen fields; while undirected data mining reveal patterns which not based on certain fields, which were defined before.

Terms "rule induction" and "association rule" can be found in data mining. The rule induction is a major technique of data mining and one of the most used of data mining in discovering knowledge in an unsupervised system. This technique is a form of data mining that is the most aligned with what people think about data mining, mining gold from a huge data. The thing that becomes the gold is interesting rules which can describe what's hidden in data and easily understood. Rule induction technique involves a great data processing, where every possible pattern can be extracted systematically from data, then be assessed in order to know when the patterns can be found again.

The data mining methodology consists of ten steps [5, 54-86]

a. Translating the business problem to a data mining problem
b. Selecting appropriate data
c. Getting to know the data
d. Creating a model set
e. Fixing problems with the data

f. Transforming data
g. Building models
h. Assessing models
i. Deploying models
j. Assessing results

3. Analytical Customer Relation Management (CRM)

CRM is a process of managing the interactions between the enterprise and its customers [6, 10]. The goal of CRM is to create and manage the dynamic environment, as efforts to improve the quality of the enterprise - its customers - interactions continuously.

There are three types of CRM, operational CRM, analytical CRM, and Collaborative CRM. The analytical CRM is a type of CRM which been used by enterprises to conduct analysis of customers behaviour. According to what's been aforementioned [4, 5] says that the analytical CRM is the usages of customers data - analyzing, modeling, and evaluating - which been purposing to support the enterprise to create profitable relations with its customers.

An analytical CRM must be a computerized system because it would handle a huge and complex data. The analytical CRM consists of all programs to analyze data about customers, in order to make faster and better decisions [2, 2].

4. Distribution Channel

A distribution channel is a path where goods been sent from producers to wholesalers and the final point, the end-users [7, 25]. The delivering products is not just a sending action, but the effectiveness and efficiency have to be considered, as said [1] Distribution is all about getting your product/service to the right people at the right time with special consideration for profit and effectiveness.

In business distribution, there are several components that should be assessed [7, 63-65]:

a. Spreading. The distributor maps the area based on some particular parameter, such as geographical environment, total population, distance, total outlets, etc. The next step is locating distribution points to serve potential outlets over there. So, the core of spreading is expanding the distribution area based on the production capacity of principals to meet the demand of that area.
b. Coverage. This means principals products are sold at every existing outlet. The prerequisite is a routine visitation to the outlets and adequate information throughout.
c. Penetration. This means infiltrating action of various products to outlets. This activity depends on the ability of sales forces. But, those must be fit to needs, wants, and hopes of customers.

The distribution does not just optimizing the spreading, but also the covering and penetrating –to – outlets ability. Spreading, coverage, and penetration (SCP) form collaboration, called network. Expanding distribution area without covering and penetrating will be useless. So, SCP would help distributing products to better products availability and faster customers searching.

5. Analysis of Current System

a. Porter Analysis

Based on the analysis using the five forces of Porter, the forces have been analyzed and categorized based on the possibility to occur and the impacts to the enterprise. This can be seen at the priority matrix table 1.

Table 1. Issues Priority Matrix

Occurrence Possibility \ Impact to Company	High	Medium	Low
High	• The bargaining power of buyers • The threat of new entrants • The competitive rivalry in industry		
Medium	• The bargaining power of supplier		
Low		• The threat of substitution products	

The conclusion of the priority matrix above:

- The competitive rivalry in industry is high because the enterprise must consider every new strategy of other enterprise.
- The bargaining power of supplier (principals) is medium because the corporation between the principal and the enterprise bounded in a legal agreement. So, the principal can not violate it by supplying to other enterprise.
- The bargaining power of buyers (outlets) is stronger than the enterprise because the number of distribution enterprise getting more, so, buyers can move from one company to another company.

- The threat of new entrants is high. This is possible due to the agreement of Asean Free Trade Area (AFTA) in pharmacy sector, which implementing through synchronizing registration. This causes more national enterprises been acquired by foreign enterprises.
- The threat of substitution products is low because traditional medicine can not replace all functionality of modern medicine. So, the modern medicine is still the fist priority of end-users.

b. SWOT Analysis

Based on SWOT analysis and IFAS EFAS calculation, the enterprise will develop the third quadrant strategy, weakness – opportunity strategy. There are two strategies:

- Cooperating with local outlets and distributors to act as sub-agents
- Maximizing the products penetration to extended channel

Between them, the second strategy is more suitable to do immediately.

c. Critical Success Factor (CSF) Analysis

To clarify the strategy, the enterprise has examining CSF and KPI, stated at the table 2.

Table 2. Critical Success Factors and Key Performance Indicators

CSF (Critical Success Factors)	KPI (Key Performance Indicators)
The location of outlets or local distributors	In radius minimal 40% from total area
The number of customers of each outlet of local distributor	Having customers, minimal 15% from the number of local population
Sales force performance	Score 76 – 90 (max.100)
The number of visitation of each salesman to each outlet	50 – 60 visits per week
The capacity of delivery and sales force fleet	All outlets can be covered with max. 150 outlets for each salesman
The number product variety for certain customer segment	40 – 50 products per outlet

d. Analysis of Current Computerized System

The architecture of information system of the enterprise can be seen as Figure 1. The operational application handles enterprise operational data. The data is emerging transforming process, so, can be kept in the data warehouse. It will be apportioned to particular data marts. The source of the information portal of enterprise and producing various reports is those data marts.

e. Problem and Information Need Analysis

Based on the analysis of the current system, esp. the implementation of the second strategy, there are three problems:

- The target of sales force is still in short-term. So, they just focus to their personal target.
- Sales force visitations depend on the condition of the traffic.
- The information about combinable products is inadequate.

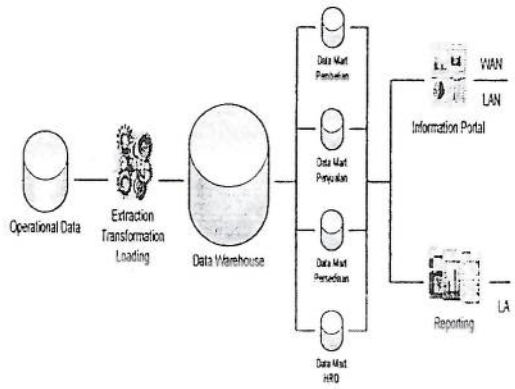


Figure 1. Current Architecture of Operational Database, Data Warehouse, and Data Mart

From those problems, only the third problem can be solved by the computerized system. The demanded information to overcome the third problem is the combination pattern of products which bought by customers, based on geographical area, customer groups, and the customer itself. It can assist the management of the enterprise to create a strategy, such as bounding campaign, promotion, etc, as a response to the customer buying behaviour in leading to the principals' products penetration.

6. Data Mining Design

In designing data mining, there are seven steps:

a. Translating the business problem to a data mining problem

To solve the business problem by using data mining, the initial phase to be done is translating the business problem to a defined problem that data mining can solve. Based on the analysis done before, the problem which enterprise facing is affinity grouping. So, the algorithm that will be used is Microsoft Association Rules.

b. Selecting appropriate data

The most appropriate source of data that available in current architecture information system and will be used is sales data mart, which is a part of the data warehouse. The architecture of the

operational database, the data warehouse, and the data mining can be seen at the Figure 2.

c. Getting to know the data

Figure 3 shows the star schema of the sales data mart that will be used in data mining. The explanation of dimension tables and fact tables of the data mart:

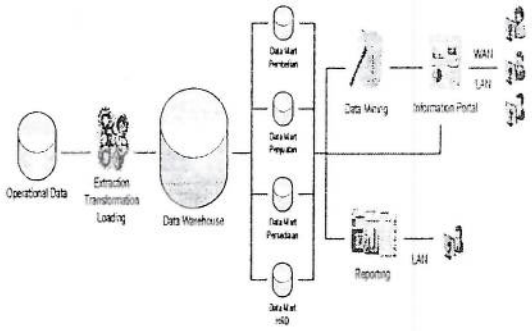


Figure 2. Proposed Architecture of Operational Database, Data Warehouse, and Data Mining

- The dimension table *DimCustomer* records the data of all enterprise's customers. One record is a snapshot of a customer's condition at particular time.
- The dimension table *DimSalesPerson* records the data of all enterprise's sales force. One record is a snapshot of a salesman's condition at particular time.
- The dimension table *DimMaterial* records the data of all products distributed by the enterprise. One record is a snapshot of a product's condition at particular time.
- The dimension table *DimSalesOffice* records the data of all enterprise's branch offices. One record is a snapshot of an office's condition at particular time.
- The dimension table *DimDate* records the data of time complement. One record represents a particular time. The dimension table *FactSales* records the data of all sales transactions. One record is a detail of a transaction.

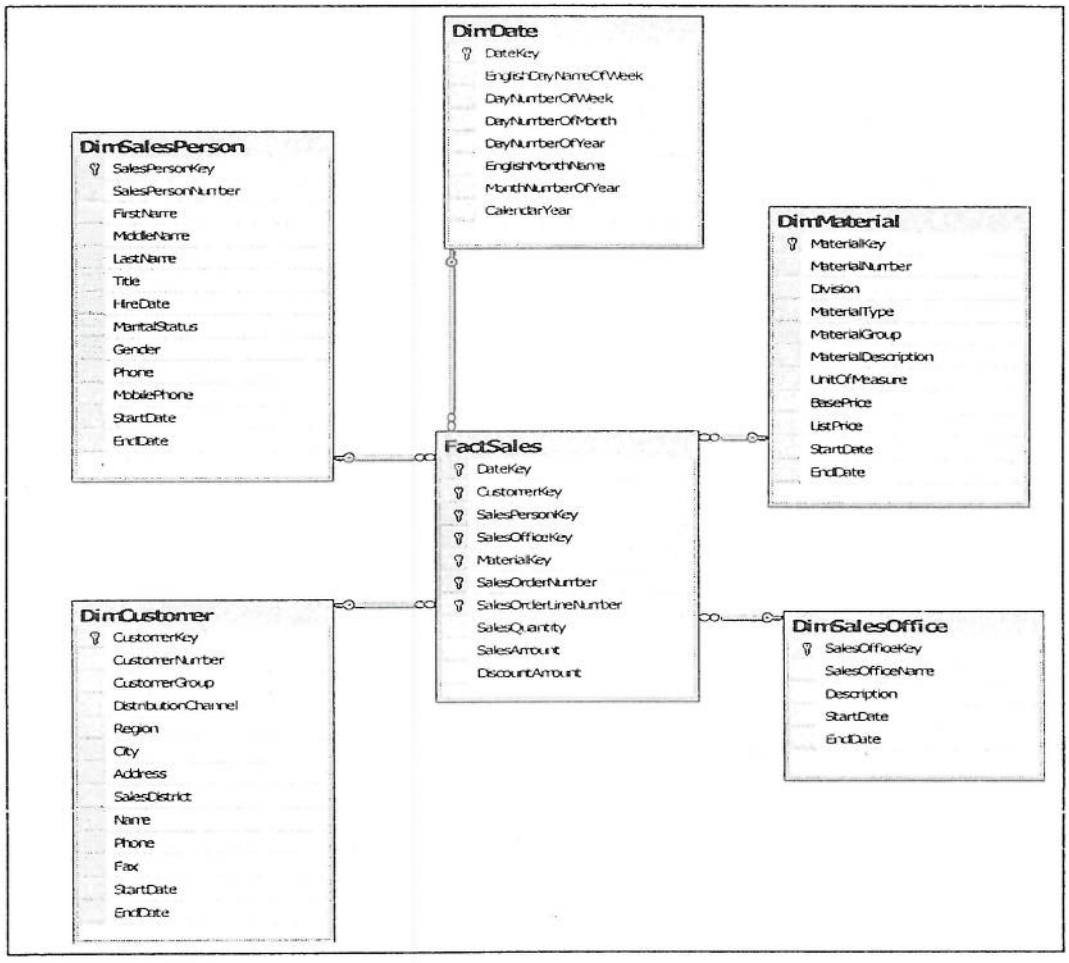


Figure 3. Star Schema of Sales Data Mart

d. Creating a model set
After understanding the data, the next phase is preparing the data to be processed.

Each type of data mining algorithm needs an input from a data source to conduct the process of model training which structured based on particular format. Formatting the data is done at this phase. Based on the chosen data mining algorithm, Microsoft Association Rules, The input data must be in a case table and nested table.

To create a format compatible data source, two data mart view must be made. Table 3 and 4 are the vNested and vCase view. The outlook of the nested table can be seen at Figure 4.

Besides constructing data based on a usable format for data mining algorithm, the data criteria must be defined for training process at particular time. The time limit must be defined afore the training process because the result of it shows the customers buying behaviour based on customer category and location at particular time

Table 3. Metadata View vNested

Field Name	Original Field	Source Table	Notes
SalesOrderNumber	SalesOrderNumber	FactSales	Sales order number
MaterialDescription	MaterialDescription	DimMaterial	Product name

Table 4. Metadata View vCase

Field Name	Original Field	Source Table	Notes
CustomerGroup	CustomerGroup	DimCustomer	Customer group number
Region	Region	DimCustomer	Customer province
CustomerName	CustomerName	DimCustomer	Customer name
SalesOrderNumber	SalesOrderNumber	FactSales	Sales order number (to connect case table with nested table)

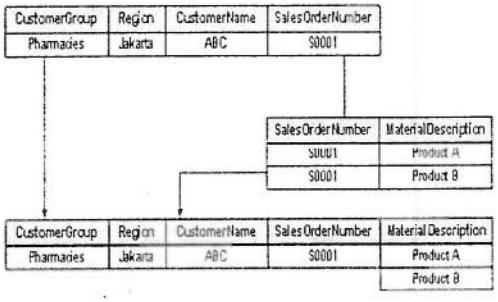


Figure 4. Outlook of nested table of designed view

e. Fixing problems with the data and Transforming data

This phase goal is to fix and transform the data necessarily in order to obtain appropriate form and quality of data to be used in data mining. Because the data is originating from data mart which a part of data warehouse (already through cleansing process), this phase is no longer needed.

f. Building models
At this phase, the model is built according to the chosen algorithm. The model will be able to show the correlation of a product to another based on the demography condition and certain customer group. Table 5 and 6 state the specification of the model structure and data mining parameter, which will be built.

Table 5. Structure Specification Designed Data Mining Model

Field Name	Key	Input	Predict	Content Type	Data Type
CustomerGroup		X		Discrete	Text
CustomerName		X		Discrete	Text
Region		X		Discrete	Text
SalesOrderNumber	X			Key	Text
MaterialDescription	X	X	X	Key	Text

Table 6. Parameter Specification of Designed Model Data Mining

Parameter Name	Parameter Value	Notes
Minimum_Support	0,03	Percentage of minimal number occurrence of itemset to support a rule
Minimum_Probability	0,4	Minimal profitability value for every rule

g. Assessing models

At this phase, a web-based application is being coded, so can be used the executives to get to know the patter that been revealed bye the model. The editor of this application is Microsoft ASP.NET and a supporting component, Dundas Map for ASP.NET.

The factors that should be considered while coding the application are the usability of the application to users and user-friendly interface.

5. User Interface

Figure 5, 6, 7, 8, 9 and 10 shows some designed pages of analytical CRM application with data mining.

6. Conclusion

The conclusion that can be made:

- The designed model data mining can reveal the pattern of customer buying behaviour that can be used by the executives to create various strategies.
- The designed model data mining can explore the data transaction of customers and produce buying prediction that can be used to anticipate customers' needs and wants.
- The phase assessing result can be executed because the produced patterns are useless due to the dummy data.
- The phase assessing can be executed because the enterprise does not do the investment action in the designed data mining.

7. Suggestions

The possible suggestions to develop the designed data mining:

- To add model to comprehend other aspects of customers besides the buying behaviour, e.g. customers segmentation, customer profitability, etc
- To fix the application architecture to minimize the response time of model training process
- To review the hardware specification along the amount of data that will be used

Figure 5. Create Model Page

Figure 6. Train Model Page

Figure 7. Customer Distribution Page

Figure 8. Customer Details Page

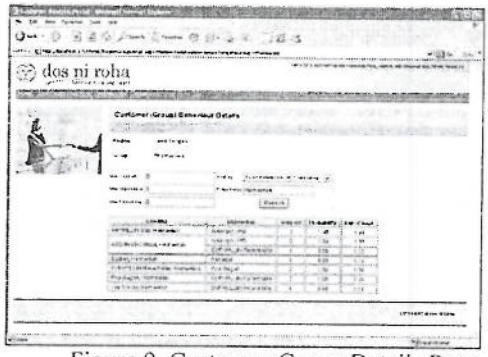


Figure 9. Customer Group Details Page

Figure 10. Prediction Page

8. References

[1] Anonym, 2008, *Distribution Channels* **http://www.bizhelp24.com/marketing/distribution-channels-4.html**

[2] Anonym, 2001, *Entering to Analytical CRM*, http://www.crmgriot.com/comp_analytical_crm.htm

[3] Anonym, 2004, *Pilihan Implementasi CRM*, http://www.ebizzasia.com/0217-2004/enterprise.0217.01.html

[4] Beck, B., Summer, J., 2001, *Data Warehousing Horizons: CRM: Not Just Operational and Collaborative*, http://www.dmreview.com/master.cfm?NavID=198&EdID=3956

[5] Berry, M. J. A., Linoff, G. S., 2004, *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*, 2nd Edition. Wiley Publishing, Inc., Indianapolis.

[6] Berson, A. Smith, S., and Thearling, K., 2000, *Building Data Mining Applications for CRM*. The McGraw-Hill Companies, Inc., U.S.A.

[7] Royan, F. M., 2004, *Sales Force: Meningkatkan Penjualan dengan Rancang Bangun Sales Force Effective*. ANDI, Yogyakarta

[8] Witten, I. H., Frank, E., 2005, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd Edition. Morgan Kaufmann Publisher, San Fransisco.